Online tools for uncovering data quality (DQ) issues in satellite-based global precipitation products

Zhong Liu[1,2] and Gil Heo[2]

[1] NASA GES DISC

[2] CSISS, George Mason University

GES – DISC
Goddard Earth Sciences
Data Information Services Center

Outline

- Introduction
- DQ challenges
- Online tools
- Examples
- Conclusion
- Future work
- Acknowledgements
- Links/information

Introduction

- Precipitation is a critical variable in the global hydrologic cycle and our daily lives (droughts, floods, diseases, etc.).

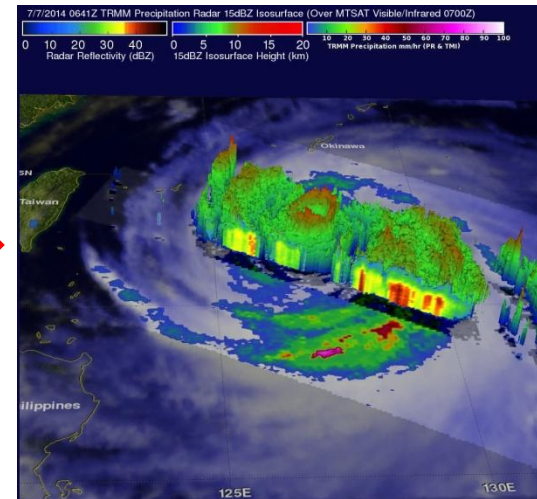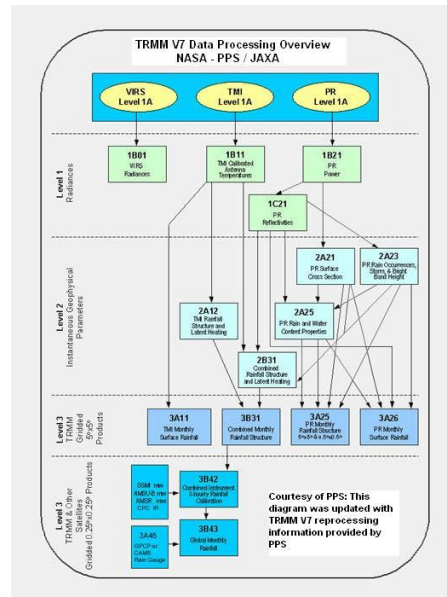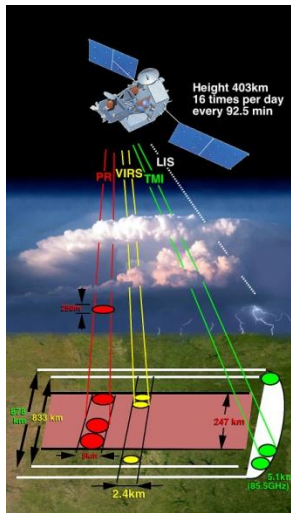Yangtze River near ChongQing during the 2010 droughts in SW China
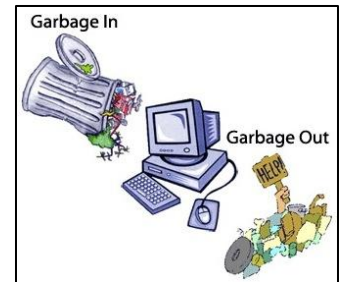

1991 Bangladesh cyclone. Fatalities: ~140,000

Introduction (Cont'd)

- Data quality (DQ) information is very important in research and applications

In a nutshell

Garbage In
Garbage Out

TRMM satellite

Algorithms

Data products

GES – DISC
Goddard Earth Sciences
Data Information Services Center

Introduction (Cont'd)

DQ in precipitation products:

- Observation issues (accuracy, spatial and temporal sampling, etc.)

- Single satellite issues (anomaly, orbit boost, etc.)

- Multi-satellite and multi-sensor issues, plus gauge (for bias correction) issues (more complicated)

- Algorithms and production issues (physics, software bugs, etc.)

Challenges

- DQ has many attributes or facets (i.e., errors, biases, systematic differences, uncertainties, benchmark, false trends, false alarm ratio, etc.)

- Sources can be complicated (measurements, environmental conditions, surface types, algorithms, etc.) and difficult to be identified especially for multi-sensor and multi-satellite products with bias correction (TMPA, IMERG, etc.)

- How to obtain DQ info fast and easily, especially quantified info in ROI? Existing parameters (random error), literature, DIY, etc.

- How to apply the knowledge in research and applications.

- Here, we focus on online systems for integration of products and parameters, visualization and analysis as well as investigation and extraction of DQ information.

Online Tools

- Giovanni – TOVAS for TRMM Versions 6 and 7 intercomparisons (TMI, PR, Multi-satellite products). Level-3 3-hourly, daily and monthly datasets. Latest: New capabilities and new datasets from different projects and missions

- Greg Leptoukh et al. – "Addressing and presenting quality of satellite data via web-based services" (aerosol datasets). Available online: http://tw.rpi.edu/media/2014/11/23/32cb/Leptoukh_ESTF2011.pdf

- Others, i.e. CSU CRDC Chris Kummerow

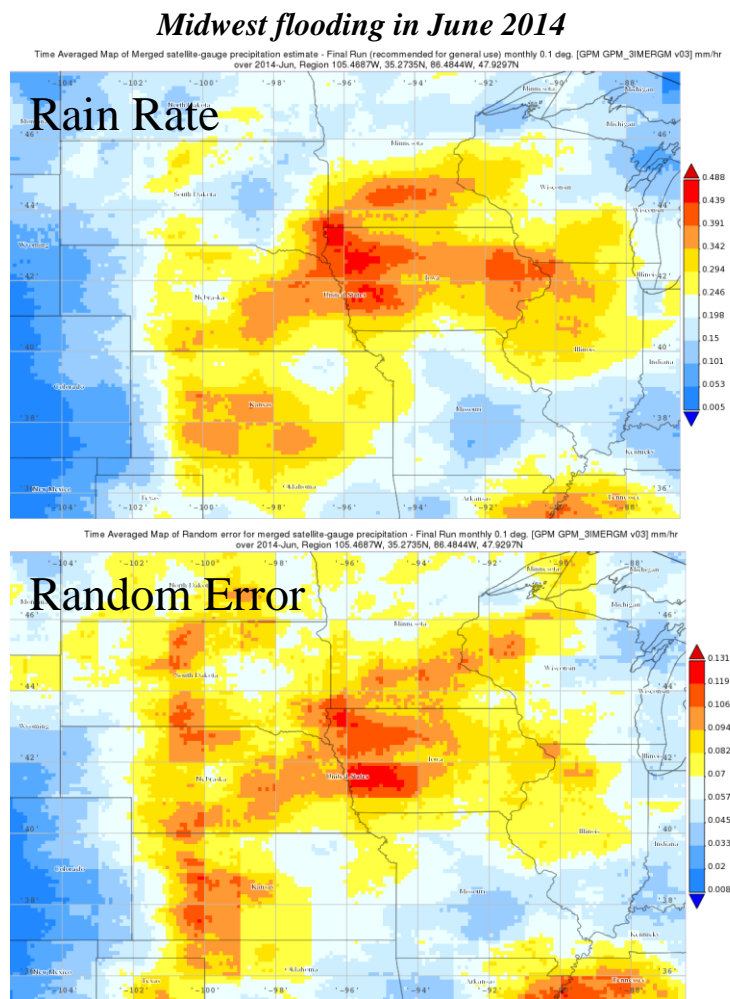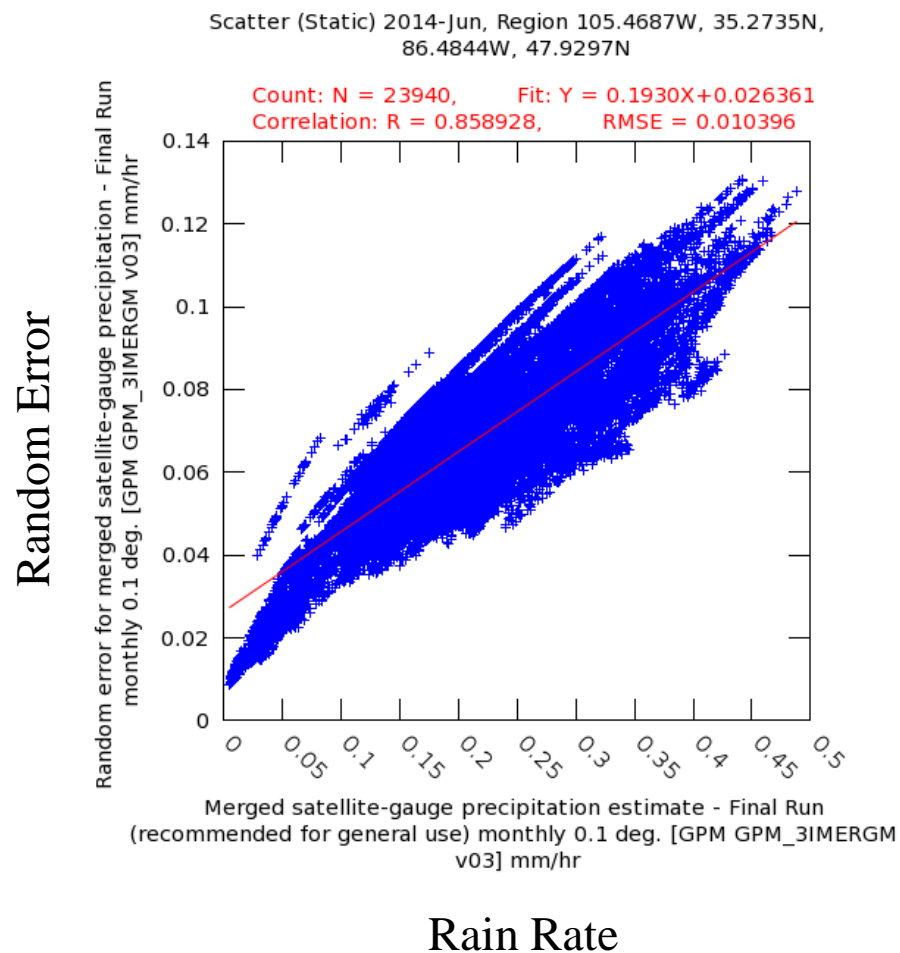# Giovanni: http://giovanni.gsfc.nasa.gov

Examples

- Existing parameters (random errors in TMPA and IMERG)

- Online visualization and analysis (TRMM PR monthly data, TMPA monthly data, and GPM IMERG half-hourly products (Early, Late and Final))

An Example: Random Error

- Given in standard products (TMPA, IMERG (monthly below))

Scatter (Static) 2014-Jun, Region 105.4687W, 35.2735N, 86.4844W, 47.9297N

Count: N = 23940, Fit: Y = 0.1930X+0.026361
Correlation: R = 0.858928, RMSE = 0.010396

Random Error

Merged satellite-gauge precipitation estimate - Final Run (recommended for general use) monthly 0.1 deg. [GPM GPM_3IMERGM v03] mm/hr

Rain Rate

Midwest flooding in June 2014

Rain Rate

Random Error

TRMM PR Monthly Product (3A25)

- TRMM orbit boost from 350 km to 403 km in Aug. 2001
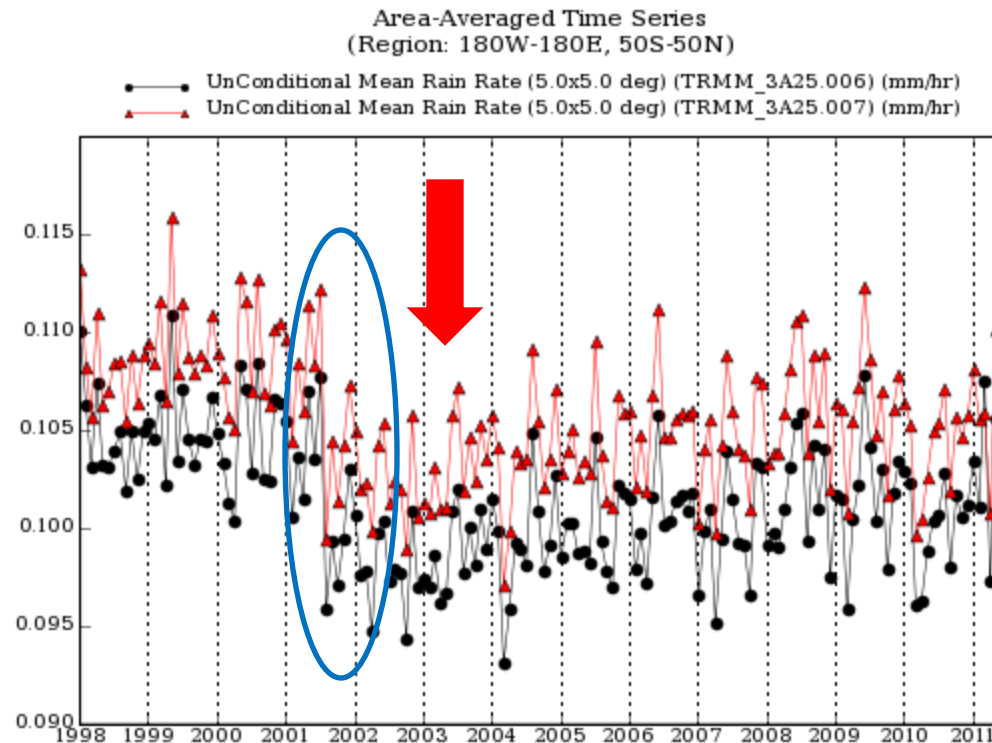
Area-Averaged Time Series
(Region: 180W-180E, 50S-50N)

TRMM PR Monthly Product (Cont'd)
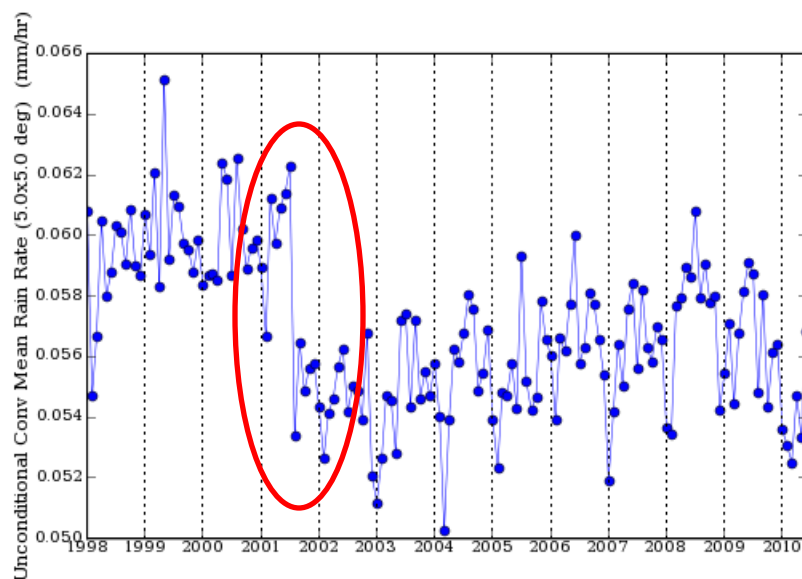
Convective (unconditional)
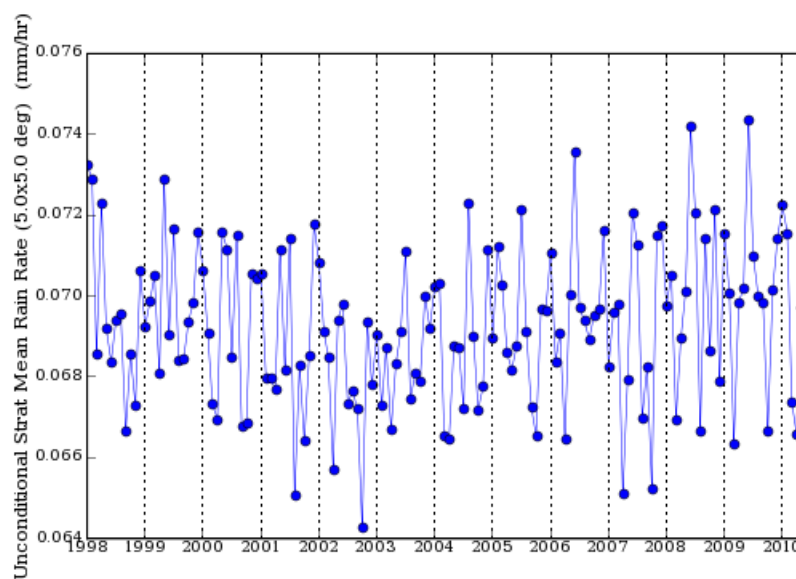
Stratiform (unconditional)

TRMM PR Monthly Product (Cont'd)

Conditional Mean Rain
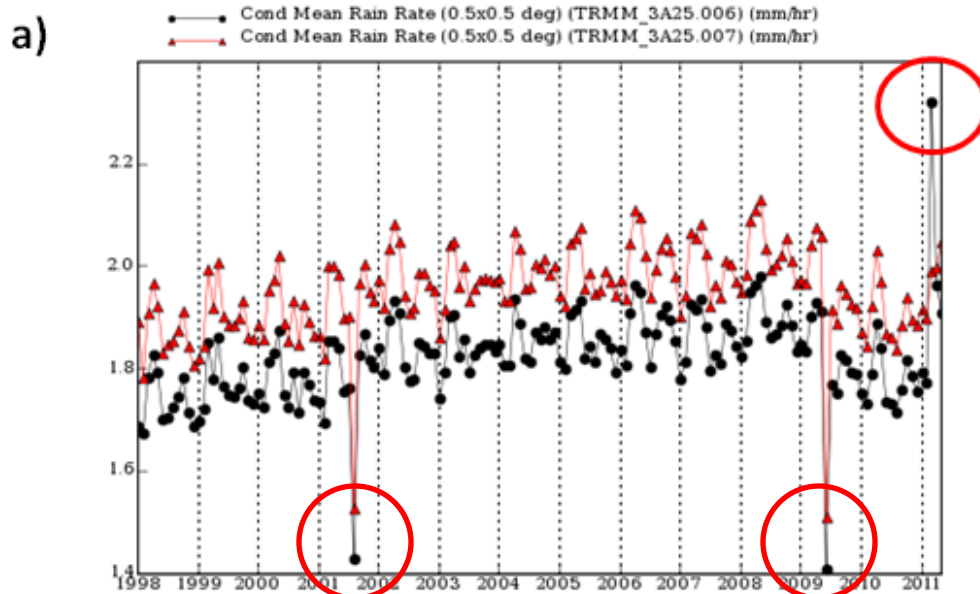
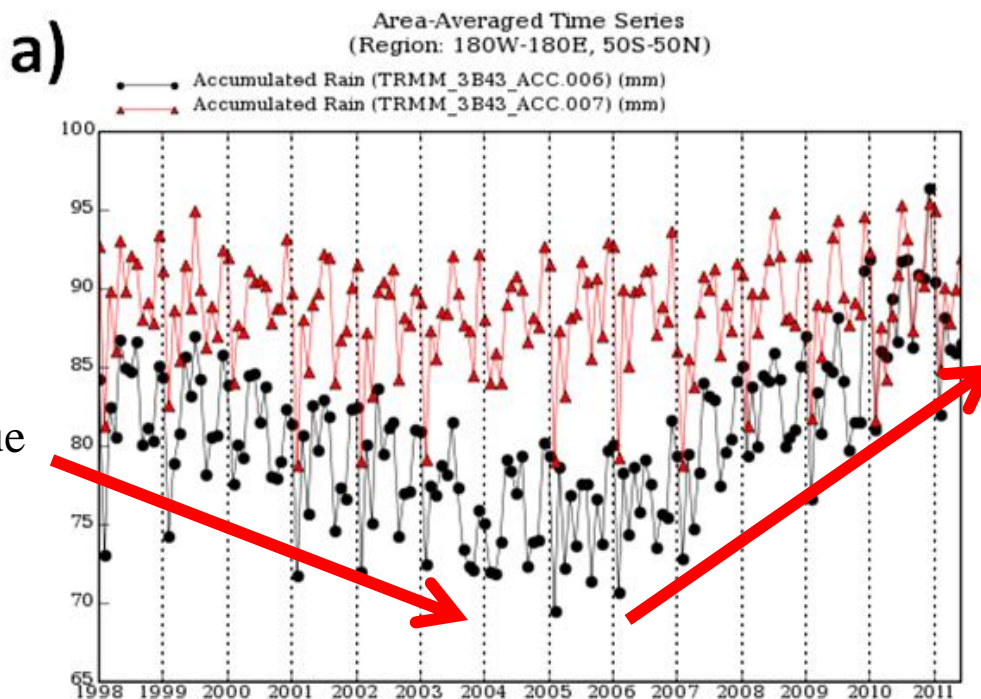Empty granules between 7th-24th 2001 due to the orbit boost

Missing data between 29th May – 18th Jun 2009 due to PR anomaly
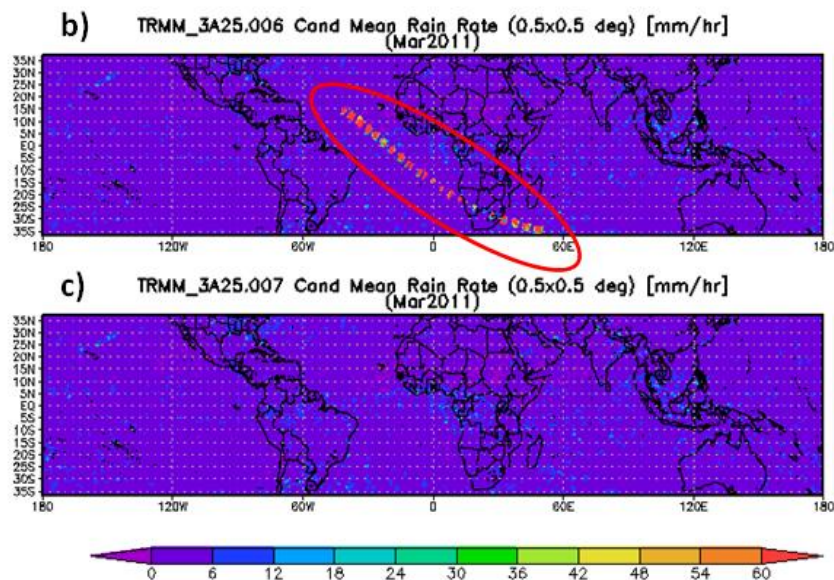
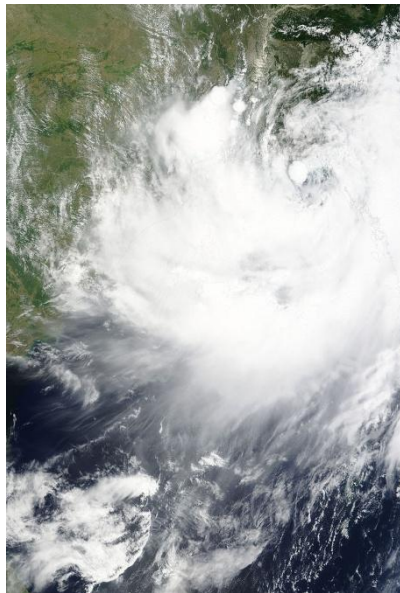# TMPA Versions 6 and 7

- 3B43 monthly product

A V-shape issue in V6

Latest Giovanni

- Inter-comparisons among Integrated Multi-satellitE Retrievals for GPM (IMERG) Early, Late and Final half-hourly products (Poster in H049)
- Myanmar Flooding in July 2015

IMERG Accumulated rainfall in mm (July 10 – July 31, 3015)

Tropical Storm Komen

Topography
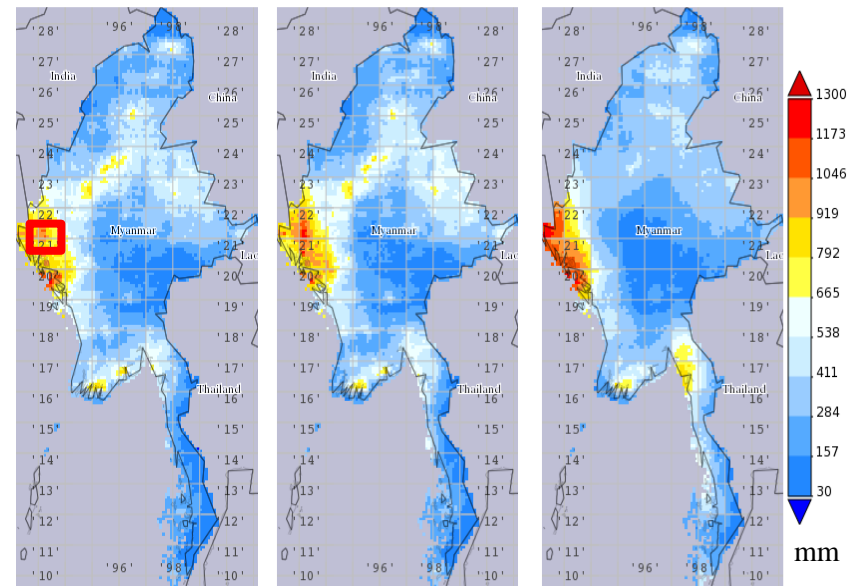
Early   **Late**   **Final**

IMERG (Cont'd)

Early (X) vs. Late (Y)　　　　Early (X) vs. Final (Y)　　　　Late (X) vs. Final (Y)

Conclusion

- Importance to have DQ information in research and applications

- Online tools (no data and software download needed) facilitate DQ investigation through data exploration and comparison such as GES DISC Giovanni

- Giovanni examples (TRMM and GPM datasets)

Future Plans

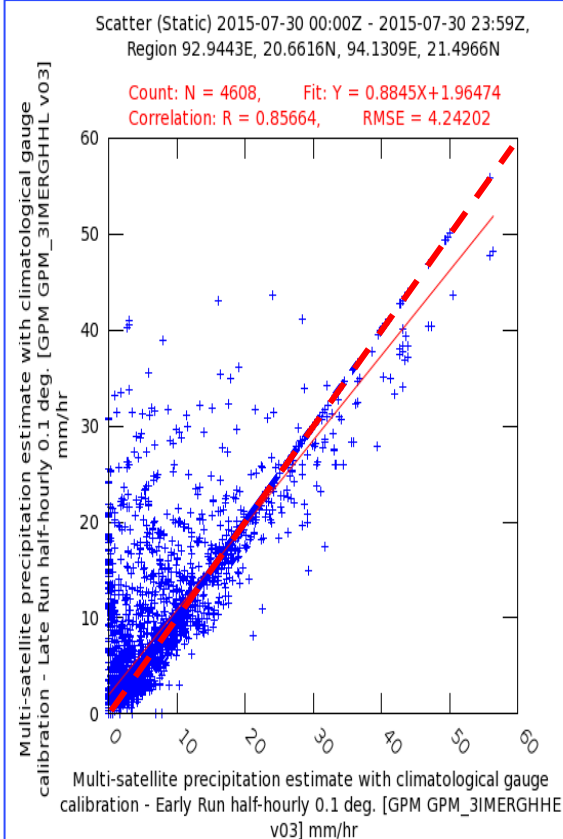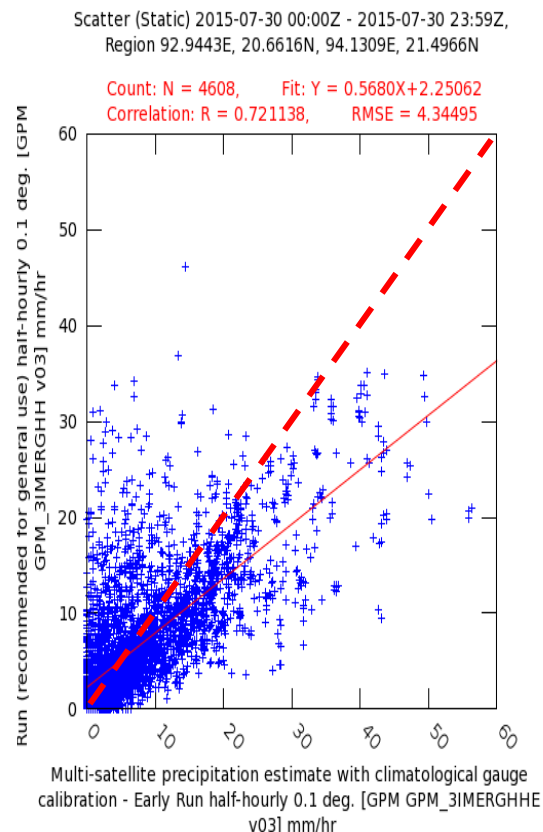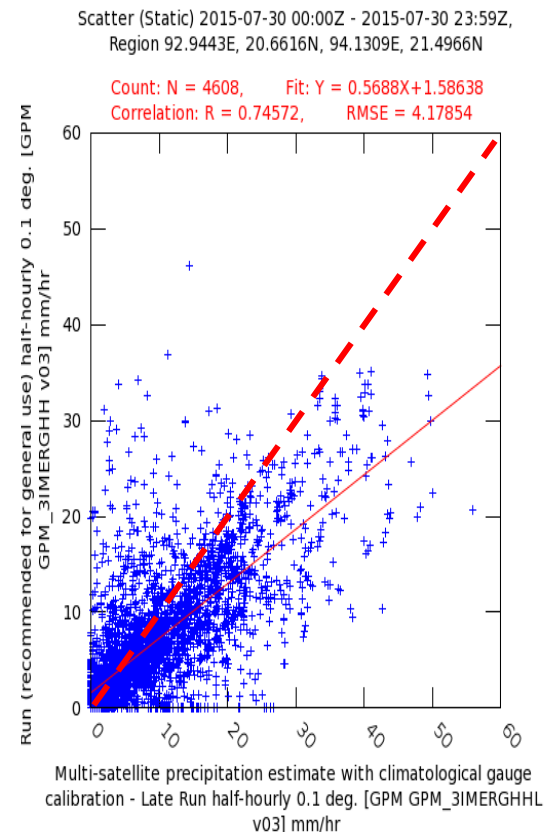Need more information and knowledge in online DQ tool development:

- Better understand DQ attributes or facets (a collection of best practices in different disciplines, for example) and algorithms for analysis

- Prototypes in several key DQ areas or disciplines (measurements, input datasets, algorithms, output datasets, etc.). A library of algorithms or toolbox? Trace-back capabilities.

- More data products. For precipitation, atmospheric and environmental conditions, orbital products (Level-1, Level-2), ancillary information (satellite, sensor anomaly, etc.), bertter grouping of similar products, etc.

- Research: error propagation (3-hourly to daily), application of knowledge, etc.

Acknowledgements

- NASA Research Opportunities in Space and Earth Science- 2010 (ROSES-2010), NNH10ZDA001N-ESDRERR, Appendix A.32: "Earth System Data Records Uncertainty Analysis"

- NASA GES DISC Giovanni Team

Information

- Giovanni: http://giovanni.gsfc.nasa.gov or Google search "NASA giovanni"
- Greg Leptoukh's ppt: http://tw.rpi.edu/media/2014/11/23/32cb/Leptoukh_ESTF2011.pdf
- My poster: *Explore GPM IMERG and Other Global Precipitation Products with GES DISC GIOVANNI* (Today in Moscone South Poster Hall H13H-1646, 13:40-1800)
- Comments and suggestions for new features: **gsfc-help-disc@lists.nasa.gov**